

УДК 004.412:519.25

Л.М. МАКАРОВА, Н.В. ПРИХОДЬКО, О.О. КУДІН

Національний університет кораблебудування ім. адм. Макарова, м. Миколаїв

ПОБУДОВА НЕЛІНІЙНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ВЕБ-ДОДАТКІВ, РЕАЛІЗОВАНИХ МОВОЮ JAVA

Розглянуто проблему отримання ефективної системи оцінки кількості строк коду та оцінювання розміру веб-додатків, реалізованих мовою Java. Побудована нелінійна регресійна модель для оцінювання розміру веб-додатків, реалізованих мовою Java, на основі одновимірного нормалізуючого перетворення Джонсона. Проведено порівняння отриманих результатів з іншими моделями.

Розраховані границі інтервалу передбачення для лінійної моделі, нелінійної моделі на основі десятичного логарифму та нелінійної моделі на основі нормалізуючого перетворення Джонсона. Усі значення нижньої границі інтервалу передбачення для нелінійних моделей більші нуля. Ширина інтервалу передбачення нелінійної регресії на основі нормалізуючого перетворення Джонсона менше, ніж для лінійної регресії, майже для всіх проектів. При порівнянні нелінійних моделей, ширина інтервалу передбачення моделі на основі нормалізуючого перетворення Джонсона менша для великих значень вихідних емпіричних даних.

Також порівняно значення коефіцієнту детермінації R^2 , середньої величини відносної похибки MMRE та рівня прогнозування PRED(0,25) для трьох побудованих регресійних моделей. Значення наведених параметрів кращі для нелінійної регресійної моделі на основі нормалізуючого перетворення Джонсона, однак прийнятні значення MMRE та PRED(0,25) (не більше 0,25 та не менше 0,75 відповідно) для нелінійної регресії з використанням одновимірного нормалізуючого перетворення Джонсона не досягнуті, що свідчить про необхідність застосування двовимірного нормалізуючого перетворення Джонсона для урахування взаємного впливу двох випадкових величин.

Ключові слова: розмір програмного забезпечення, нелінійна регресія, нормалізуюче перетворення Джонсона, Java, веб-додаток.

Л.Н. МАКАРОВА, Н.В. ПРИХОДЬКО, О.А. КУДИН

Национальный университет кораблестроения им. адм. Макарова, г. Николаев

ПОСТРОЕНИЕ НЕЛИНЕЙНОЙ РЕГРЕССИОННОЙ МОДЕЛИ ДЛЯ ОЦЕНИВАНИЯ РАЗМЕРА ВЕБ-ПРИЛОЖЕНИЙ, РЕАЛИЗОВАННЫХ НА ЯЗЫКЕ JAVA

Рассмотрена проблема получения эффективной системы оценки количества строк кода и оценивания размера веб-приложений, реализованных на языке Java. Построена нелинейная регрессионная модель для оценивания размера веб-приложений, реализованных на языке Java, на основе одномерного нормализующего преобразования Джонсона. Проведено сравнение полученных результатов с другими моделями.

Рассчитаны границы интервала предсказания для линейной модели, нелинейной модели на основе десятичного логарифма и нелинейной модели на основе нормализующего преобразования Джонсона. Все значения нижней границы интервала предсказания для нелинейных моделей больше нуля. Ширина интервала предсказания нелинейной регрессии на основе нормализующего преобразования Джонсона меньше, чем для линейной регрессии, почти для всех проектов. При сравнении нелинейных моделей, ширина интервала предсказания модели на основе нормализующего преобразования Джонсона меньше для больших значений исходных эмпирических данных.

Также произведено сравнение значений коэффициента детерминации R^2 , средней величины относительной погрешности MMRE и уровня прогнозирования PRED (0,25) для трех построенных регрессионных моделей. Значения приведенных параметров лучше для нелинейной регрессионной модели на основе нормализующего преобразования Джонсона, однако приемлемые значения MMRE и PRED (0,25) (не более 0,25 и не менее 0,75 соответственно) для нелинейной регрессии с использованием одномерного нормализующего преобразования Джонсона не достигнуты, что свидетельствует о необходимости применения двумерного нормализующего преобразования Джонсона для учета взаимного влияния двух случайных величин.

Ключевые слова: размер программного обеспечения, нелинейная регрессия, нормализующее преобразование Джонсона, Java, веб-приложение.

L.M. MAKAROVA, N.V. PRYKHODKO, O.O. KUDIN
Admiral Makarov National University of Shipbuilding, Mykolaiv

CONSTRUCTING THE NON-LINEAR REGRESSION MODEL FOR SIZE ESTIMATION OF WEB-APPLICATIONS IMPLEMENTED IN JAVA

The problem of obtaining an effective system for estimating the number of lines of code and size estimating of web applications implemented in Java was considered. The non-linear regression model for size estimation of Java web-applications were constructed on the basis of the univariate Johnson normalizing transformation. Obtained results were compared with other models.

The bounds of a prediction interval for the linear model, the nonlinear model based on the decimal logarithm, and the nonlinear model based on the Johnson normalizing transformation, was calculated. All values of the lower bound of the prediction interval for nonlinear models are more than zero. The width of a prediction interval of nonlinear regression based on the Johnson normalizing transformation is less than for linear regression for almost all projects. In case comparing nonlinear models, the width of a prediction interval of model based on the Johnson normalizing transformation is less for large values of the initial empirical data.

Comparison of the values of coefficient of determination R^2 , mean magnitude of the relative error MMRE and percentage of prediction PRED (0,25) for the three constructed regression models was performed too. The values of all parameters are better for the non-linear regression model based on the Johnson normalization transformation. However, acceptable values of MMRE and PRED (0,25) (no more than 0,25 and no less than 0,75, respectively) for non-linear regression model with using the univariate Johnson normalizing transformation are not achieved, that indicates the need to use the bivariate Johnson normalizing transformation to consider a mutual impact of two random variables.

Keywords: software size, non-linear regression, Johnson normalizing transformation, Java, web-application.

Постановка проблеми

Галузь інформаційних технологій на сьогоднішній день є однією з найбільш успішних і перспективніших на ринку праці, і розробка програмного забезпечення (ПЗ) на замовлення стає сьогодні вже не просто вдалою ідеєю, а необхідністю. Все більша кількість компаній прагне замовити розробку ПЗ. Це пов'язано з прагненням оптимізувати процеси складання звітності, управління ресурсами і проведення різних операцій. І великі корпорації, і порівняно невеликі, ще молоді компанії, які тільки беруть курс на розширення, замислюються про необхідність автоматизувати як окремі бізнес-процеси, так і роботу в цілому.

Розмір ПЗ являє собою одну з найцікавіших, але у той же час найскладніших, метрик ПЗ, який використовується в різних моделях для прогнозування вартості, зусиль, ресурсів, необхідних для розробки та впровадження ПЗ [1].

Отримання ефективної системи оцінки кількості строк коду в даний час є важливим завданням, що вимагає удосконалення існуючих методів. Адже саме ефективність оцінки розміру ПЗ може стати відправною точкою для успіху або невдачі проекту на ранньому етапі розробки. Саме тому дану проблему можна вважати актуальною.

Аналіз останніх досліджень і публікацій

Розмір ПЗ є однією з найвагоміших метрик в управлінні процесом розробки ПЗ. Доведено, що розмір ПЗ корелює з витратами, зусиллями та ресурсами, необхідними на його розробку. В наш час існує досить поширена проблема в галузі інформаційних технологій: не існує єдиного вимірювання, набору метрик та показників для оцінки розмірів ПЗ.

Моделі оцінювання розміру ПЗ поділяються на п'ять категорій: аналогові; регресійні; моделі на основі експертних оцінок; моделі, які базуються на функціональних точках; параметричні моделі [2].

Існують різні методи для оцінки розміру ПЗ, які використовуються сьогодні. Більшість з них походять від методу аналізу функціональних точок (FPA). Інший підхід полягає в тому, щоб провести функціональне вимірювання та виразити функціональність у кількості, що представляє розмір. Інші методи визначення розміру ПЗ включають оцінку на основі варіантів використання (Use Case). Але історично найпоширенішою та найбільш вживаною методологією визначення розміру ПЗ є підрахунок кількості рядків, написаних у вихідному коді програми.

Крім того, всесвітньо відомою є модель СОСОМО – модель регресії, яка заснована на кількості рядків коду (LOC). Ця процедурна модель оцінки витрат для програмних проектів часто використовується для надійного прогнозування різних параметрів, пов'язаних з проектом, таких, як розмір, зусилля, витрати, час та якість, які необхідні для розробки та впровадження ПЗ. Інформацію, отриману у результаті оцінки розміру ПЗ, можна використати для прогнозування зусиль розробки ПЗ за такими моделями, як СОСОМО 81, СОСОМО II та СОСОМО 2000 [3]. Але більшість з них стають непридатними до використання через відсутність даних, ресурсів або експертних навичок в цій галузі.

Згідно з [4], із метрик, отриманих на основі діаграм класів можна побудувати нелінійну регресійну модель для багатовимірних негаусівських даних для оцінки розміру ПЗ, використовуючи нормалізуюче перетворення Джонсона, та отримати непогані результати у порівнянні з іншими регресійними моделями. Суть застосованої методики полягає в наступному. Спочатку потрібно обрати перетворення, за допомогою якого здійснити перехід від вихідних негаусівських випадкових величин (ВВ) до гаусівських ВВ. Після цього для отриманих нормалізованих ВВ побудувати лінійну регресійну модель. Далі за допомогою зворотного перетворення перейти до нелінійної регресійної моделі вихідних негаусівських ВВ [5].

Формулювання мети дослідження

Метою роботи є побудова нелінійної регресійної моделі для оцінювання розміру веб-додатків, реалізованих мовою Java, з використанням нормалізуючого перетворення Джонсона.

Для досягнення поставленої мети потрібно вирішити наступні задачі:

- проаналізувати існуючі моделі оцінювання розміру ПЗ, порівняти їх;
- побудувати нелінійну регресійну модель для оцінювання розміру веб-додатків, реалізованих мовою Java;
- дослідити джерела з відкритим вихідним кодом та визначити веб-додатки, реалізовані мовою Java, які можуть бути використані для перевірки побудованої моделі;
- побудувати діаграми класів та отримати необхідні метрики з кожного проекту;
- перевірити вихідні емпіричні дані на викиди;
- нормалізувати отримані емпіричні дані, використовуючи нормалізуюче перетворення Джонсона;
- побудувати лінійну регресійну модель для нормалізованих даних;
- перейти від лінійної регресії до нелінійної та побудувати нелінійну регресійну модель для вихідних даних.

Викладення основного матеріалу дослідження

У роботі [4] для оцінки розміру ПЗ побудовано нелінійну регресійну модель на основі багатовимірного нормалізуючого перетворення Джонсона. Але побудова багатовимірної моделі на основі декількох метрик достатньо складна та затратна в плані зусиль та часу, необхідних для її реалізації. Вона може стати непридатною до використання через відсутність даних, ресурсів або експертних навичок в цій галузі. Саме тому в якості першого наближення для оцінювання кількості рядків коду веб-додатків, реалізованих мовою Java, буде доцільно побудувати нелінійну регресійну модель на основі двох метрик: кількості рядків коду та загальної кількості класів.

Для нормалізації негаусівських ВВ можуть бути використані перетворення на основі десяткового або натурального логарифму, перетворення Вох-Сох, перетворення Джонсона та інші. У даній роботі в якості нормалізуючого перетворення буде використовуватись одновимірне чотирихпараметричне перетворення Джонсона тому, що воно дає кращі результати в порівнянні з іншими відомими перетвореннями [5].

При обробці експериментальних даних часто виникає необхідність апроксимувати їх. Апроксимація на основі сімей розподілів Джонсона – це універсальний вид апроксимації, заснований на такому перетворенні $h(x)$ вихідної ВВ X , яке дозволить розглядати результат перетворення як ВВ, розподілену за нормальним законом. Одним з позитивних аспектів даного підходу є те, що значення емпіричної функції розподілу ВВ X обчислюються як значення функції нормального розподілу [6].

Перетворення Джонсона в загальному випадку має вигляд:

$$z = \gamma + \eta h(x, \varphi, \lambda); -\infty < \gamma < \infty; \eta > 0; -\infty < \varphi < \infty; \lambda > 0, \quad (1)$$

де z – нормована нормально розподілена ВВ; $\gamma, \eta, \varphi, \lambda$ – параметри перетворення; x – ВВ, яка нормалізується; h – функція певної сім'ї: $h_1(x, \varphi, \lambda) = \ln(\tilde{x}), x > \varphi$; $h_2(x, \varphi, \lambda) = \ln\left(\frac{\tilde{x}}{1 - \tilde{x}}\right), \varphi < x < \varphi + \lambda$;

$$h_3(x, \varphi, \lambda) = \text{Arsh}(\tilde{x}), -\infty \leq x \leq +\infty.$$

Сім'ї функцій h_1 відповідає логарифмічно нормальний розподіл S_L Джонсона, сім'ї функцій h_2 відповідає сім'я розподілів S_B Джонсона, сім'ї функцій h_3 відповідає сім'я розподілів S_U Джонсона, $\tilde{x} = \frac{x - \varphi}{\lambda}$.

Перетворення (1) має зворотне перетворення:

$$x = \varphi + \lambda h^{-1}(z, \gamma, \eta); -\infty < \gamma < \infty; \eta > 0; -\infty < \varphi < \infty; \lambda > 0, \quad (2)$$

де h^{-1} – функція певної сім'ї: $h_1^{-1}(z, \gamma, \eta) = e^{\zeta}$; $h_2^{-1}(z, \gamma, \eta) = \frac{1}{1+e^{-\zeta}}$; $h_3^{-1}(z, \gamma, \eta) = \frac{e^{\zeta} - e^{-\zeta}}{2}$.

Функція h_1^{-1} – для сім'ї S_L Джонсона, функція h_2^{-1} – для сім'ї S_B Джонсона, функція h_3^{-1} – для сім'ї S_U Джонсона, $\zeta = \frac{z-\gamma}{\eta}$.

Конкретна сім'я розподілу Джонсона вибирається виходячи із значень квадрата асиметрії A^2 і ексцесу ε вихідної вибірки [7]:

$$\varepsilon(A^2) = 3,59 \cdot 10^{-6} A^8 - 4,8805 \cdot 10^{-4} A^6 + 4,1655 \cdot 10^{-2} A^4 + 1,8203 A^2 + 2,9658. \quad (3)$$

Значення невідомих параметрів розподілу можна знайти за допомогою методу максимальної правдоподібності [8].

Перевірку відповідності перетворених вибірок нормальному розподілу можна виконати за допомогою критеріїв згоди, наприклад, χ^2 Пірсона або Колмогорова – Смирнова [9].

Для оцінювання розміру веб-додатків, реалізованих мовою Java, однією з основних задач є побудова відповідної математичної моделі, у нашому випадку, це буде регресійна модель [10, 11], яке буде нелінійною [12].

Загальний вигляд лінійної регресійної моделі може бути представлений у вигляді рівняння [13]:

$$z_y = b_1 z_x + b_0 + \varepsilon, \quad (4)$$

де b_1, b_0 – коефіцієнти лінійної регресії, які знаходяться методом найменших квадратів, ε – гаусівська ВВ, що визначає залишки (нев'язки) та визначається як $\varepsilon \sim N(0,1)$.

Для перевірки адекватності лінійного рівняння регресії використаємо коефіцієнт детермінації R^2 [14].

Далі для лінійної регресії будемо інтервал передбачення традиційним способом із застосуванням t -розподілу Стюдента [5, 15].

Для побудови нелінійної регресійної моделі використаємо вже побудовану лінійну регресійну модель (4) та зворотнє нормалізуюче перетворення Джонсона (2):

$$y = \frac{e^c (\lambda_y + \phi_y) + \phi_y}{1 + e^c} + \varepsilon, \quad (5)$$

$$\text{де: } c = \frac{1}{\eta_y} \cdot \left(b_1 \left[\gamma_x + \eta_x \ln \left(\frac{x - \phi_x}{\lambda_x + \phi_x - x} \right) \right] + b_0 - \gamma_y \right).$$

Рівень якості побудованого нелінійного рівняння регресії оцінимо за допомогою середньої величини відносної похибки MMRE та рівня прогнозування PRED(0,25) [8].

$(1-\alpha)\%$ інтервал передбачення нелінійної регресії можна побудувати, використовуючи лінійну регресійну модель (4), t -розподіл Стюдента та зворотнє нормалізуюче перетворення Джонсона (2):

$$y = \frac{e^{k_1} (\lambda_y + \phi_y) + \phi_y}{1 + e^{k_1}}, \quad (6)$$

$$\text{де: } k_1 = \frac{1}{\eta_y} \cdot \left(b_1 \cdot z_x + b_0 - \gamma_y \pm t_{(\alpha/2, n-2)} \cdot S_{z_y} \cdot \sqrt{1 + \frac{1}{n} + \frac{(z_x - \bar{z}_x)^2}{\sum_{i=1}^n (z_{xi} - \bar{z}_x)^2}} \right), \quad z_x = \gamma_x + \eta_x \ln \left(\frac{x - \phi_x}{\lambda_x + \phi_x - x} \right).$$

Перевірку даних на викиди виконаємо з використанням еліпсу передбачення згідно з методикою, викладеною у [16, 17]. Точки, які знаходяться за межами еліпсу, вважаються викидами, і ці значення видаляються з набору даних.

Перевірку запропонованої методики здійснимо, використовуючи емпіричні дані 33 веб-додатків, реалізованих мовою Java, з відкритим вихідним кодом. Кожен з веб-додатків було проаналізовано та побудовано до нього діаграму класів і набір метрик, у результаті чого отримано двовимірний набір

даних, де X – загальна кількість класів, Y – число рядків коду веб-додатку в KLOC. Отримані емпіричні дані для вибірок X та Y не відповідають нормальному закону розподілу.

Перевірка емпіричних даних на викиди та побудова еліпсу передбачення виявили три викиди, два з яких були отримані на першій ітерації (див. рис. 1), третій – на другій ітерації, які були видалені з вихідного набору даних.

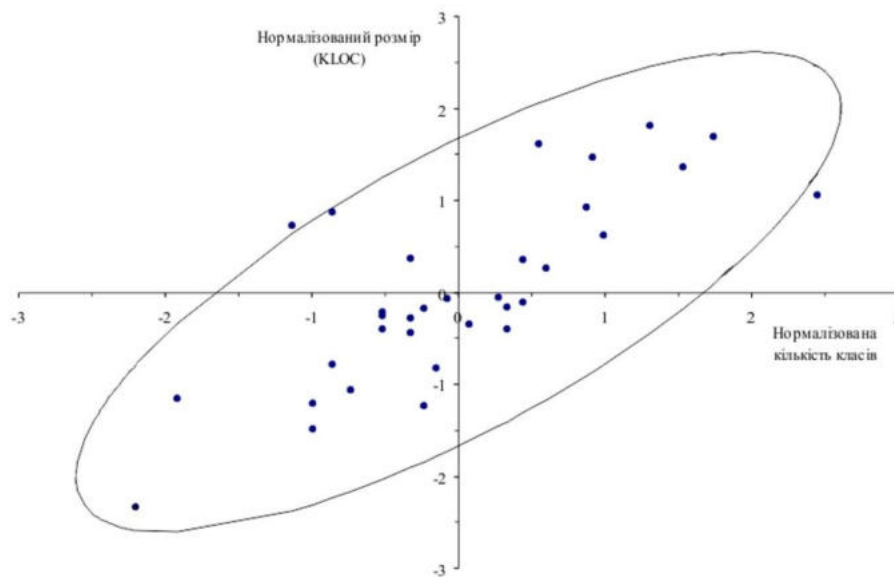


Рис. 1. Еліпс передбачення вихідних емпіричних даних 33 веб-додатків, реалізованих мовою Java

Для нормалізації емпіричних даних виходячи із значень квадрата асиметрії A^2 і ексцесу ε вихідних вибірок ($A_X^2=4,90$, $\varepsilon_X=9,49$, $A_Y^2=3,40$, $\varepsilon_Y=5,90$) була обрана сім'я розподілів Джонсона S_B згідно з (3). Нормалізацію даних для вибірок X та Y проводимо з використанням нормалізуючого перетворення Джонсона (1). Значення невідомих параметрів γ , η , ϕ , λ для перетворення Джонсона знаходимо методом максимальної правдоподібності.

Параметри розподілу вибірки X : $\hat{\gamma}_X = 9,3571$, $\hat{\eta}_X = 1,5013$, $\hat{\phi}_X = 0,4560$, $\hat{\lambda}_X = 10742,70$.

Параметри розподілу вибірки Y : $\hat{\gamma}_Y = 1,5704$, $\hat{\eta}_Y = 0,7006$, $\hat{\phi}_Y = 0,2220$, $\hat{\lambda}_Y = 12,8680$.

Якість нормалізації перевіримо із застосуванням критерію згоди χ^2 Пірсона: $\chi^2_X=1,82 < \chi^2_{кр}=7,81$, $\chi^2_Y=4,45 < \chi^2_{кр}=7,81$. З довірчою ймовірністю 0,95 гіпотеза про відповідність перетворених вибірок Z_X та Z_Y нормальному закону розподілу приймається.

Вихідні та нормалізовані дані 30 веб-додатків, реалізованих мовою Java, наведено у табл. 1.

Побудуємо лінійну регресійну модель для нормалізованих даних згідно з (4), коефіцієнти знайдемо за допомогою методу найменших квадратів: $b_1=0,8773$, $b_0=0,00$, $z_y = 0,8773 \cdot z_x + \varepsilon$. Коефіцієнт детермінації $R^2=0,6147$. Отже, отримане лінійне рівняння регресії можна вважати прийнятним.

Далі будуємо інтервал передбачення лінійної регресії, який разом із самим рівнянням та нормалізованими вихідними даними наведено на рис. 2.

Переходимо до вихідних емпіричних даних. На основі лінійної регресійної моделі та зворотнього нормалізуючого перетворення будуємо нелінійну регресійну модель згідно з (5):

$$y = \frac{13,09 \cdot e^c + 0,222}{1 + e^c} + \varepsilon, \text{ де } c = 9,4756 + 1,88 \cdot \ln\left(\frac{x - 0,456}{10743,16 - x}\right).$$

Таблиця 1

Вихідні та нормалізовані дані 30 веб-додатків, реалізованих мовою Java

№ проекту	X	Y	Z _X	Z _Y	№ проекту	X	Y	Z _X	Z _Y
1	18	1,320	-0,2745	-0,0913	16	15	1,050	-0,5565	-0,3049
2	17	1,020	-0,3628	-0,3325	17	57	6,200	1,4879	1,4709
3	12	0,739	-0,9037	-0,6528	18	103	4,609	2,3880	1,1086
4	38	6,878	0,8704	1,6188	19	20	1,457	-0,1122	-0,0008
5	19	0,716	-0,1912	-0,6859	20	11	0,516	-1,0399	-1,0607
6	49	9,444	1,2577	2,2205	21	26	1,057	0,2906	-0,2986
7	28	2,229	0,4040	0,3875	22	11	0,423	-1,0399	-1,3323
8	5	0,261	-2,3044	-2,4899	23	65	8,470	1,6877	1,9764
9	17	1,177	-0,3628	-0,1975	24	18	0,507	-0,2745	-1,0830
10	22	1,109	0,0344	-0,2533	25	37	3,999	0,8298	0,9551
11	6	0,543	-2,0057	-0,9977	26	28	1,395	0,4040	-0,0406
12	15	1,267	-0,5565	-0,1291	27	13	0,586	-0,7789	-0,9072
13	15	1,210	-0,5565	-0,1718	28	26	1,327	0,2906	-0,0865
14	31	2,034	0,5597	0,3034	29	40	2,930	0,9486	0,6441
15	17	2,263	-0,3628	0,4015	30	25	1,488	0,2305	0,0185

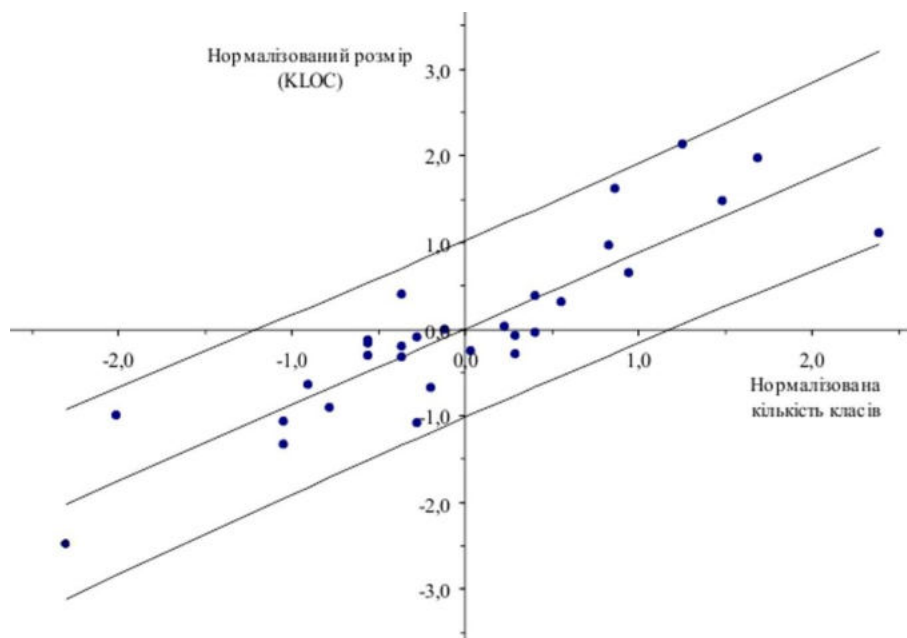


Рис. 2. Лінійне рівняння регресії та інтервал передбачення для 30 веб-додатків, реалізованих мовою Java

Далі будемо рівняння верхньої та нижньої границь інтервалу передбачення нелінійної регресії згідно з (6): $y = \frac{13,09 \cdot e^{k_1} + 0,222}{1 + e^{k_1}}$ де:

$$k_1 = 1,4273 \cdot \left(0,8773 \cdot z_x - 1,5704 \pm t_{(\alpha/2, n-2)} \cdot S_{z_y} \cdot \sqrt{1 + \frac{1}{30} + \frac{(z_x - \bar{z}_x)^2}{\sum_{i=1}^n (z_{xi} - \bar{z}_x)^2}} \right),$$

$z_x = 9,3571 + 1,5013 \cdot \ln\left(\frac{x - 0,456}{10743,16 - x}\right)$, які разом із нелінійним рівнянням регресії та вихідними емпіричними даними наведено на рис. 3.

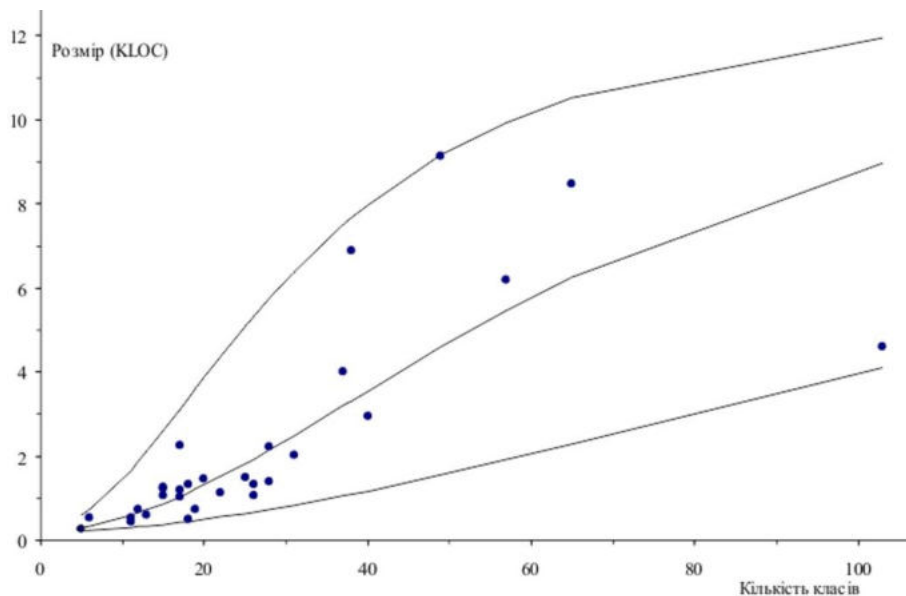


Рис. 3. Нелінійне рівняння регресії та інтервал передбачення для 30 веб-додатків, реалізованих мовою Java

Для порівняння моделі (5) з іншими моделями побудовані такі регресійні моделі на основі вихідних емпіричних даних (див. табл. 1): лінійна регресійна модель без виконання нормалізації (в припущенні про нормальність вихідних емпіричних даних) $y = 0,0893 \cdot x - 0,1194 + \varepsilon$ та нелінійна регресійна модель на основі десяткового логарифмічного перетворення $y = 10^{\varepsilon - 1,469} \cdot x^{1,2266}$. Границі для інтервалів передбачення цих регресійних моделей наведено у табл. 2.

Таблиця 2

Границі інтервалів передбачення регресійних моделей

№ про-екту	Лінійна модель		Нелінійна модель на основі десяткового логарифму		Нелінійна модель на основі нормалізуючого перетворення Джонсона	
	нижня границя	верхня границя	нижня границя	верхня границя	нижня границя	верхня границя
1	2	3	4	5	6	7
1	-1,99	4,97	0,47	2,96	0,44	3,42
2	-2,09	4,88	0,44	2,76	0,42	3,16
3	-2,55	4,45	0,28	1,82	0,32	1,93
4	-0,21	6,76	1,16	7,48	1,07	7,75
5	-1,90	5,06	0,50	3,16	0,46	3,67
6	0,72	7,80	1,56	10,34	1,53	9,23
7	-1,09	5,85	0,80	5,10	0,72	5,84
8	-3,21	3,86	0,09	0,66	0,24	0,59
9	-2,09	4,88	0,44	2,76	0,42	3,16
10	-1,63	5,32	0,60	3,78	0,54	4,43
11	-3,11	3,95	0,12	0,81	0,25	0,73
12	-2,27	4,71	0,37	2,38	0,37	2,66
13	-2,27	4,71	0,37	2,38	0,37	2,66
14	-0,82	6,12	0,91	5,79	0,82	6,48
15	-2,09	4,88	0,44	2,76	0,42	3,16
16	-2,27	4,71	0,37	2,38	0,37	2,66
17	1,37	8,57	1,86	12,57	1,89	10,00
18	4,87	13,29	3,66	27,30	4,05	11,98
19	-1,81	5,14	0,53	3,37	0,49	3,93
20	-2,64	4,37	0,25	1,64	0,30	1,70
21	-1,27	5,67	0,73	4,65	0,66	5,39
22	-2,64	4,37	0,25	1,64	0,30	1,70

Продовження таблиці 2

1	2	3	4	5	6	7
23	2,01	9,36	2,17	14,90	2,26	10,58
24	-1,99	4,97	0,47	2,96	0,44	3,42
25	-0,30	6,67	1,12	7,23	1,04	7,59
26	-1,09	5,85	0,80	5,10	0,72	5,84
27	-2,46	4,54	0,31	2,00	0,34	2,17
28	-1,27	5,67	0,73	4,65	0,66	5,39
29	-0,04	6,95	1,23	7,98	1,15	8,06
30	-1,36	5,58	0,70	4,43	0,63	5,16

Як видно із табл. 2, нижня границя інтервалу передбачення лінійної моделі має від'ємні значення для 26 проєктів. Усі значення нижньої границі інтервалу передбачення для нелінійних моделей більші нуля. Ширина інтервалу передбачення нелінійної регресії на основі нормалізуючого перетворення Джонсона менше, ніж для лінійної регресії, для 27 проєктів. При порівнянні нелінійних моделей ширина інтервалу передбачення на основі нормалізуючого перетворення Джонсона менше для великих значень вихідних емпіричних даних.

Також порівняємо значення коефіцієнту детермінації R^2 , середньої величини відносної похибки MMRE та рівня прогнозування PRED(0,25) для трьох побудованих регресійних моделей, які наведено у табл. 3.

Таблиця 3

Значення R^2 , MMRE та PRED(0,25)

Параметр	Лінійна модель	Нелінійна модель на основі десятичного логарифму	Нелінійна модель на основі нормалізуючого перетворення Джонсона
R^2	0,5497	0,4845	0,6147
MMRE	0,4894	0,3575	0,3477
PRED(0,25)	0,3667	0,4667	0,5000

Як видно з табл. 3, значення наведених параметрів кращі для нелінійної регресійної моделі на основі нормалізуючого перетворення Джонсона, однак прийнятні значення MMRE та PRED(0,25) (не більше 0,25 та не менше 0,75 відповідно) для нелінійної регресії з використанням одновимірного нормалізуючого перетворення Джонсона не досягнуті, що свідчить про необхідність застосування двовимірного нормалізуючого перетворення Джонсона для урахування взаємного впливу двох ВВ.

Висновки

Побудовано нелінійну регресійну модель для оцінювання розміру веб-додатків, реалізованих мовою Java, на основі одновимірного нормалізуючого перетворення Джонсона. Виконано порівняння отриманих результатів з іншими моделями. У подальшому планується використання багатовимірного нормалізуючого перетворення Джонсона для нормалізації емпіричних даних щодо оцінювання розміру веб-додатків, реалізованих мовою Java, та розробка відповідного програмного забезпечення для автоматизації процесу розрахунків.

Список використаної літератури

1. Briand L.C. Property Based Software Engineering Measurement / L.C. Briand, S. Morasca, V.R. Basili // IEEE Transaction on Software Engineering. – 2009. – Vol. 22, no. 1. – p. 68–86.
2. Briand L.C. Encyclopedia of Software Engineering / L.C. Briand, I. Wiecek – John Wiley & Sons, Inc., New Jersey, 2002. – 1584 p.
3. Tan H.B.K. Estimating LOC for information systems from their conceptual data models / H.B.K. Tan, Y. Zhao, H. Zhang // in Proceedings of the 28th International Conference on Software Engineering (ICSE '06), May 20–28, 2006, Shanghai, China. – P. 321 – 330.
4. Prykhodko S. Estimating the software size of open-source PHP-based systems using non-linear regression analysis / S. Prykhodko, N. Prykhodko, L. Makarova // in Proceedings of International Conference on Advanced Computer Information Technologies (ACIT 2018), June 1–3, 2018, Ceske Budejovice, Czech Republic. – P. 199 – 202.
5. Приходько С.Б. Доверительный интервал нелинейной регрессии времени восстановления работоспособности устройств терминальной сети / С.Б. Приходько, Л.Н. Макарова // Восточно-европейский журнал передовых технологий. Математика и кибернетика – прикладные аспекты. – 2014. – Т. 3/4 (69). – С. 26 – 31.

6. Бостанджиян В.А. Распределение Пирсона, Джонсона, Вейбулла и обратное нормальное. Оценивание их параметров / В.А. Бостанджиян – Черноголовка: Редакционно-издательский отдел ИПХФ РАН, 2009. – 240 с.
7. Приходько С.Б. Аналитическая зависимость для выбора семейства распределений Джонсона / С.Б. Приходько, Л.Н. Макарова, А.С. Приходько // Проблеми інформаційних технологій. – 2016. – №02 (020). – С. 105 – 110.
8. Prykhodko S.B. Building The Non-Linear Regression Equations On The Basis Of Multivariate Normalizing Transformations / S. Prykhodko, N. Prykhodko, L. Makarova // Proceedings of First International Conference on System Analysis & Intelligent Computing (SAIC-2018), Kyiv, Ukraine, October 08–12, 2018. – P. 48 – 52.
9. Вентцель Е.С. Теория вероятностей: Учеб. для вузов / Е.С. Вентцель – М.: Высш. шк., 1999. – 576 с.
10. Грешилов А.А. Математические методы построения прогнозов / А.А. Грешилов, В.А. Стакун, А.А. Стакун – М.: Радио и связь, 1997. – 112 с.
11. Демиденко Е.З. Линейная и нелинейная регрессии / Е.З. Демиденко – М.: Финансы и статистика, 1981. – 302 с.
12. Bates D.M. Nonlinear Regression Analysis and Its Applications / D.M. Bates, D.G. Watts – John Wiley & Sons, Inc., New Jersey, 1988. – 384 p.
13. Prykhodko N.V. Constructing the non-linear regression models on the basis of multivariate normalizing transformations / N.V. Prykhodko, S.B. Prykhodko // Electronic modeling. – 2018. – Vol. 40, No. 6. – P. 101 – 110.
14. Магнус Я.Р. Эконометрика. Начальный курс: Учеб. – 6-е изд., перераб. и доп. / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. – М.: Дело, 2004. – 576 с.
15. Приходько С.Б. Інформаційна технологія прогнозування відмов в обслуговуванні пристроїв термінальної мережі / С.Б. Приходько, Л.М. Макарова // Проблеми інформаційних технологій. – 2015. – №01 (017). – С. 187 – 194.
16. Chew V. Confidence, prediction and tolerance regions for the multivariate normal distribution / V. Chew // Journal of the American Statistical Association. – 1966. – Vol. 61, Issue 315. – P. 605 – 617.
17. Prykhodko S.B. Detecting bivariate outliers on the basis of normalizing transformations for non-Gaussian data / S.B. Prykhodko, N.V. Prykhodko, L.M. Makarova, O.O. Kudin, T.G. Smykodub, A.S. Prykhodko // Advanced Information Systems and Technologies: proceedings of the V international scientific conference, Sumy, May 17-19 2017 Edited by S.I. Protsenko, V.V. Shendryk. – Sumy: Sumy State University, 2017. – P. 95 – 97.